STANFORD UNIVERSITY SCHOOL OF MEDICINE
STANFORD MEDICAL CENTER
300 PASTEUR DRIVE, PALO ALTO, CALIFORNIA

DEPARTMENT OF GENETICS

August 25, 1966

DAvenport 1-1200
Ext. 5052

Jane Wilcox, Sc.D., Division of Research Grants, National Institutes of Health, Bethesda, Maryland

Dear Doctor Wilcox:

I am eager to respond to your query for more details on various features of our proposal. Since different people have been involved in your office and on the study section during the evolution of this work is easy to see why some sections would be puzzling.

As to the title, the history of the project is documented in our applications for HD-00045-01 and -01S1 respectively. Our original request was for a rather modest effort centered on the inmate population of Pacific State Hospital, Pomona, California. We called on some census population sample data for a control base and discovered this would be only barely adequate. Then we were approached by a number of interested workers both at census and at NIH whether we would take responsibility ourselves for a larger collaborative effort to generate the Census "Child Spacing Report". We agreed to do this, and it is the main brunt of the effort now funded.

Our fundamental interest in this file is to analyze the variance of fertility. As Professor Crow pointed out this represents the limits of the "opportunity for natural selection in man." But gross fertility is not an adequate measure, and we aim to add at least some of the necessary corrections by which to estimate the "Mathusian parameter", the kinetic constant of human increase. The most important of these corrections is the maternal age at birth. More generally, we hope to find expressions with the least number of independent parameters that can describe the distributions of kindreds classified by each of the pertinent socio-economic categories given in the Census data. This would be the equivalent of a mathematical model of the population, down to the detail of age at marriage, and intervals to and between the children. We feel that such models are absolutely indispensible to predict the characterizations of the population at various epochs. The Pomona kindreds could, for example, then be contrasted with the control base in a more meaningful way giving some assay of perturbed fertility as a prognostic stigma and as a consequence of a mentally retarded birth.

The principle theoretical issue is then "what is the simplest mathematical model that will account for the population(s) described by the census?" Several distributions have been suggested in the literature, but there has been no comparable effort to fit the parameters.

Published reports could go a long way but they are never sufficiently cross tabulated on the relevant variables. Furthermore as we rather painfully discovered, a surprising proportion of the records, particularly in some of the minority categories, are incomplete or inaccurate. It has therefore, been

necessary to use a number of procedures to estimate the values of the missing variables. This is perhaps the most objectionable , precisely in the Child Spacing Study since the parameter of actual date of birth of the allocated individual becomes quite important. In our research plan, pages 3 and 6-12, we have given a fairly detailed account of the procedure we intend to follow in handling these data.

At the time your letter was mailed we were still in some doubt as to the disposition of our application for the ACME computer facility at Stanford Medical School. Since that time the proposal has been approved and funded. On that basis, we have arranged for a September 30, 1966, shipment of an IBM 360-50 Computer System. A great deal of advance work has been done preliminary to setting up this system and our present time table fits quite closely to our original schedule: namely to be operational by January 1. We will have more experimental access to the system even before then. Through some fortunate happenings we already have access to another time-shared IBM 360-50 system operated by the Allen-Babcock-Company in Palo Alto. We therefore have the basic facilities to begin working immediately in this framework. An IBM 2741 Communications terminal for this purpose is in my office at the present time.

I am enclosing some additional detailed material on the ACME system. Our plan for this set up was greeted with some skepticism but as far as I can tell this has been rather largely alleviated on the basis of more detailed examination of the technical proposal. Indeed several other computer centers have expressed deep interest in the system that is being set up here.

Until a couple of days ago, I was rather discouraged about the necessity of meeting previously established deadlines for the completion of the Child Spacing reports. This would have been barely, perhaps not quite, possible within the budget previously outlined together with the administrative supplement that we have discussed over the telephone. However, through conversations with Mr. Beresford at the Bureau of the Census, we have been able to establish a considerable prolongation of the deadlines for completion and this has the very considerable advantage that we will be able to do the bulk of the computing on the new ACME system. The main advantage of this is that resources that would have been used in a deadend program on the existing 7090 can now be applied to setting up the systems for continuous monitoring of the tabulations via the time-sharing system. Since the ACME system is primarily dedicated to the acquisition of experimental laboratory data and then processing, the census tabulations will have only a second priority although this will still be rather more favorable than what is usually regarded as "background". Except during the busiest parts of the day, it is then quite realistic that requests for tabulations of data stored on tape or in the data cell can be initiated from a typewriter console with an expected response within the attention span of the user. Besides this, the interactive typewriter arrangement, judging from my own experiences(with rather complicated LISP programs on the Q32 remote time-sharing system at System Development Corporation) leads me to predict an increased efficiency in programming by a factor of at least ten and possibly as much as a hundred in the time required between setting up the specifications for a program and actually
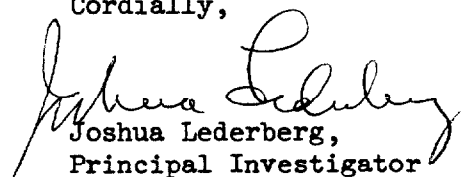
having it operational. Our experience on this score has been just as fantastic as the numbers just recorded would imply.

In order to meet the exigencies of our contract with the Bureau of the Census, we have had to forego some substantial equipment items present in our HD-0045-02. But I will leave to you for later negotiation and substantiation the rearrangements of the budget of 03 that would be appropriate to compensate for this. The overall budget as previously stated, is in fact, quite modest and we will possibly be a little short handed with respect to programing. However, since we will be making some use of the system programming effort of the ACME system and in view of the efficiencies that I just mentioned, this may not prove to be so constraining after all.

The availability of data rather than our fundamental concern for the mentally retarded does indeed limit that part of our work and it does not call on a large proportion of the funds. I should comment that the clustering of twin births mentioned on page 5 has proven to be a statistical artifact. We made the perhaps classic blunder of ignoring the self-evident correlation of births as between the twins; and when we analyze the data in terms of numbers of confinements the significance of the clustering disappears. We have also reviewed the time sequence of the occurrence of Downs' syndrome from the Pomona data and find no evidence whatsoever of the clustering that has been claimed by Collman &. Stoller and by others for these events. In fact, under the aegis of the project, our own statistical analysis of what has been reported here leaves us very sceptical of the reality of the claims that haveebeen forwarded so far for the epidemicity of this disease. We are looking somewhat more closely at the anomolous distribution of birth weight in the Pomona population and have been somewhat surprised to find that the decrement of birth weight is probably not fully accounted for by prematurity and furthermore that it applies approximately uniformally to every category of IQ. This has led us to look a little more closely at the correlation of birth weight and academic performance of children in the Palo Alto area. The hypothesis being tested is that children in extreme birth weight categories, particularly the very high birth weight, may show little peturbation of _mean_ performance but a considerable increase in the _dispersion_ of of intellectual performance. This may not be too surprising since the very fact of an anomolous birth weight suggests some developmental pecularity and there would be no a priori basis to assume that this pecularity would always operate in the same way or in the same direction with respect to intellect. This study would merely illustrate one of the many incidental subprojects that are inevitably generated by the opportunity to examine statistical data of the sort we are collecting. They typify the kind of effort which is only possible when all the statistical details are available and which are all too readily obscured when only mean values are furnished.

Cordially,

Joshua Lederberg,
Principal Investigator

JL/gem